# Regularity lemma—statement, regular pairs and their properties

Zdeněk Dvořák

September 14, 2015

Many "extremal" problems are simple when restricted to random graphs. For example, let us fix a real number $p$ such that $0 \leq p \leq 1$. What is the smallest possible number of triangles in an $n$-vertex graph with at least $p\binom{n}{2}$ edges? This is a fairly difficult problem, which was only recently completely solved by Razborov. However, if the graph is random (in the sense that each pair of vertices forms an edge independently with probability at least $p$), then clearly the graph will contain at least roughly $p^3\binom{n}{3}$ triangles.

Regularity lemma is an useful tool which essentially tells that every sufficiently large graph can be approximated by a bounded number of "random-like" subgraphs. Before we can state this result precisely, we need to introduce a number of definitions.

For a graph $G$ and disjoint non-empty sets $A, B \subset V(G)$,

- let $e(A, B)$ denote the number of edges of $G$ with one end in $A$ and the other end in $B$, and
- let $d(A, B) = \frac{e(A,B)}{|A||B|}$.

Note that $d(A, B)$ is the density of edges between $A$ and $B$—if we choose $a \in A$ and $b \in B$ at random, the probability that $ab \in E(G)$ is $d(A, B)$.

We now need a suitable notion of a "random-like" graph. It suffices to require that the edge density in every sufficiently large subgraph is close to the edge density of the whole graph.

**Definition 1.** *Let $G$ be a graph and let $\delta, \varepsilon > 0$ be real numbers. A pair $(A, B)$ of disjoint non-empty sets $A, B \subset V(G)$ is* $(\delta, \varepsilon)$-regular *if*

- *all $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \delta|A|$ and $|B'| \geq \delta|B|$ satisfy*

$$|d(A', B') - d(A, B)| \leq \varepsilon.$$

*We say that the pair is* $\varepsilon$-regular *if it is* $(\varepsilon, \varepsilon)$-regular.

Let us argue that this really describes random graphs well enough. We will need a basic result from the probability theory.

**Theorem 1** (Chernoff inequality)**.** *Let $0 \le p \le 1$ be a real number. Let $X_1$, ..., $X_n$ be independent random variables, taking value 1 with probability $p$ and value 0 with probability $1-p$. Let $X = \sum_{i=1}^{n} X_i$. Then for every $t \ge 0$,*

$$Prob(|X - pn| > t) < 2e^{-\frac{t^2}{2(pn+t/3)}}.$$

**Lemma 2.** *Let $0 \le p \le 1$ and $\varepsilon > 0$ be real numbers. There exists an integer $n_0$ such that the following holds. Let $G$ be a random graph such that each pair of vertices forms an edge independently with probability $p$. Let $A, B \subset V(G)$ be disjoint, such that $|A| = |B| \ge n_0$. Then $(A, B)$ is an $\varepsilon$-regular pair with probability at least $1 - \varepsilon$.*

*Proof.* Let $|A| = |B| = N$. Let $A' \subseteq A$ and $B' \subseteq B$ be sets of size at least $\varepsilon N$. By Theorem 1,

$$\text{Prob}(|e(A', B') - p|A'||B'|| > \varepsilon|A'||B'|/2) < 2e^{-\frac{\varepsilon^2}{8(p+\varepsilon/6)}|A'||B'|} \le 2e^{-\frac{\varepsilon^4}{8(p+\varepsilon/6)}N^2},$$

and equivalently, letting $\kappa = \frac{\varepsilon^4}{8(p+\varepsilon/6)}$,

$$\text{Prob}(|d(A', B') - p| > \varepsilon/2) < 2e^{-\kappa N^2}.$$

Note that there are at most $4^N$ pairs of subsets of $A$ and $B$, and thus the probability that there exist sets $A' \subseteq A$ and $B' \subseteq B$ of size at least $\varepsilon N$ such that $|d(A', B') - p| > \varepsilon/2$ is less than

$$4^N 2e^{-\kappa N^2} = e^{(\log 4)N + \log 2 - \kappa N^2} \le \varepsilon/2$$

for $N$ large enough.

In particular, $\text{Prob}(|d(A, B) - p| > \varepsilon/2) \le \varepsilon/2$ for $N$ large enough. Hence, with probability at least $1 - \varepsilon$, we have $|d(A, B) - p| \le \varepsilon/2$ and $|d(A', B') - p| \le \varepsilon/2$ for every $A' \subseteq A$ and $B' \subseteq B$ of size at least $\varepsilon N$. It follows that $|d(A', B') - d(A, B)| \le \varepsilon$ for all such sets $A'$ and $B'$. □

# 1 Properties of regular pairs

Let us now show that regularity implies some of the properties expected of a random graph. Firstly, almost all degrees are close to those expected according to the edge density.

**Lemma 3.** *Let $G$ be a graph and let $(A, B)$ be a $(\delta, \varepsilon)$-regular pair in $G$ for some $0 < \delta, \varepsilon \leq 1$. Then*

- *the number of vertices of $A$ with more than $(d(A, B) + \varepsilon)|B|$ neighbors in $B$ is less than $\delta|A|$, and*
- *the number of vertices of $A$ with less than $(d(A, B) - \varepsilon)|B|$ neighbors in $B$ is less than $\delta|A|$.*

*Proof.* Let $A_1$ and $A_2$ be the sets of vertices of $A$ with more than $(d(A, B) + \varepsilon)|B|$ and less than $(d(A, B) - \varepsilon)|B|$ neigbors in $B$, respectively. Suppose for a contradiction that $|A_i| \geq \delta|A|$ for some $i \in \{1, 2\}$. Let $A' = A_i$ and $B' = B$. By the $(\delta, \varepsilon)$-regularity, we have

$$|d(A', B') - d(A, B)| \leq \varepsilon.$$

However,

$$\begin{aligned} |d(A', B') - d(A, B)| &= \frac{|e(A', B) - d(A, B)|A'||B||}{|A'||B|} \\ &= \frac{\left|\sum_{v \in A'}(\deg_B(v) - d(A, B)|B|)\right|}{|A'||B|} > \frac{\varepsilon|B||A'|}{|A'||B|} = \varepsilon, \end{aligned}$$

which is a contradiction. □

In fact, we can get the following stronger claim.

**Lemma 4.** *Let $G$ be a graph and let $(A, B)$ be a $(\delta, \varepsilon)$-regular pair in $G$ for some $0 < \delta, \varepsilon \leq 1$. Let $B' \subseteq B$ have size at least $\delta|B|$. Then*

- *the number of vertices of $A$ with more than $(d(A, B) + \varepsilon)|B'|$ neighbors in $B'$ is less than $\delta|A|$, and*
- *the number of vertices of $A$ with less than $(d(A, B) - \varepsilon)|B'|$ neighbors in $B'$ is less than $\delta|A|$.*

*Proof.* Let $A_1$ and $A_2$ be the sets of vertices of $A$ with more than $(d(A, B) + 2\varepsilon)|B'|$ and less than $(d(A, B) - \varepsilon)|B'|$ neighbors in $B'$, respectively. Suppose for the contradiction that $|A_i| \geq \delta|A|$ for some $i \in \{1, 2\}$. Let $A' = A_i$. By the $(\delta, \varepsilon)$-regularity, we have

$$|d(A', B') - d(A, B)| \leq \varepsilon.$$

However,

$$\begin{aligned}|d(A',B') - d(A,B)| &= \frac{|e(A',B') - d(A,B)|A'||B'||}{|A'||B'|} \\ &= \frac{\left|\sum_{v\in A'}(\deg_{B'}(v) - d(A,B)|B'|)\right|}{|A'||B'|} \\ &> \frac{\varepsilon|B'||A'|}{|A'||B'|} = \varepsilon,\end{aligned}$$

which is a contradiction. □

We can also find roughly the expected number of more complicated subgraphs in a regular pair. For example, in a pair $(A, B)$ with density $p$, we would expect roughly $p^4|A|^2|B|^2$ 4-cycles. Here, let us give just a corresponding lower bound on the number of 4-cycles in a regular pair.

**Lemma 5.** *Let $G$ be a graph and let $(A, B)$ be a $(\delta, \varepsilon)$-regular pair in $G$ for some $0 < \delta, \varepsilon \leq 1$, with $|A| = |B| = n$ and $d(A, B) \geq \delta + \varepsilon$. The number of 4-cycles $v_1v_2v_3v_4$ of $G$ with $v_1, v_3 \in A$ and $v_2, v_4 \in B$ is at least*

$$\left[(1-\delta)^2(d(A,B) - \varepsilon)^4 - 2/n\right] n^4.$$

*Proof.* Let $A_0$ be the set of vertices of $A$ with less than $(d(A, B) - \varepsilon)n$ neighbors in $B$. By Lemma 3, $|A_0| < \delta n$.

We will count only the 4-cycles with $v_1 \in A \setminus A_0$. Let $B_{v_1}$ be the set of neighbors of $v_1$ in $B$. Then $|B_{v_1}| \geq (d(A, B) - \varepsilon)n \geq \delta n$. Let $A_{v_1}$ be the set of vertices of $A$ with less than $(d(A, B) - \varepsilon)|B_{v_1}|$ neighbors in $B_{v_1}$. By Lemma 4, $|A_{v_1}| < \delta n$.

Let us now estimate the number $c$ of choices of $v_1 \in A\setminus A_0$ and $v_3 \in A\setminus A_{v_1}$, and their common neigbors $v_2, v_4 \in B$. The number of common neighbors of $v_1$ and $v_3$ is at least $(d(A, B) - \varepsilon)|B_{v_1}| \geq (d(A, B) - \varepsilon)^2 n$. Since $|A \setminus A_0| \geq (1 - \delta)n$ and $|A \setminus A_{v_1}| \geq (1 - \delta)n$, we have $c \geq (1 - \delta)^2(d(A, B) - \varepsilon)^4 n^4$.

Not all choices counted by $c$ are 4-cycles, since we can have $v_1 = v_3$ or $v_2 = v_4$. However, the number of such "degenerate" 4-cycles is at most $2n^3$. Therefore, the number of all 4-cycles $v_1v_2v_3v_4$ of $G$ with $v_1, v_3 \in A$ and $v_2, v_4 \in B$ is at least

$$c - 2n^3 \geq \left[(1-\delta)^2(d(A,B) - \varepsilon)^4 - 2/n\right] n^4.$$

□

Finally, we need to argue that different regular pairs play nice with each other. For example, if $(A, B)$, $(B, C)$ and $(A, C)$ are "random-like", we would expect there to be roughly $d(A, B)d(B, C)d(A, C)|A||B||C|$ triangles with one vertex in each of $A$, $B$, and $C$. Again, we only prove a lower bound.

**Lemma 6.** *Let $G$ be a graph and let $(A, B)$, $(B, C)$ and $(A, C)$ be $(\delta, \varepsilon)$-regular pairs in $G$ for some $0 < \delta, \varepsilon \leq 1/2$, with $|A| = |B| = |C| = n$ and $d(A, B), d(B, C), d(A, C) \geq \delta + \varepsilon$. The number of triangles $v_1v_2v_3$ of $G$ with $v_1 \in A$, $v_2 \in B$, and $v_3 \in C$ is at least*

$$(1 - 2\delta)(d(B, C) - \varepsilon)(d(A, B) - \varepsilon)(d(A, C) - \varepsilon)n^3.$$

*Proof.* Let $A_0$ be the set of vertices of $A$ with less than $(d(A, B) - \varepsilon)n$ neighbors in $B$, and let $A_1$ be the set of vertices of $A$ with less than $(d(A, C) - \varepsilon)n$ neighbors in $C$. By Lemma 3, $|A_0|, |A_1| < \delta n$.

We will count only the triangles with $v_1 \in A \setminus (A_0 \cup A_1)$. Let $B_{v_1}$ and $C_{v_1}$ be the sets of neighbors of $v_1$ in $B$ and $C$, respectively. Then $|B_{v_1}| \geq (d(A, B) - \varepsilon)n \geq \delta n$ and $|C_{v_1}| \geq (d(A, C) - \varepsilon)n \geq \delta n$. By the $(\delta, \varepsilon)$-regularity of $(B, C)$, we have $d(B_{v_1}, C_{v_1}) \geq d(B, C) - \varepsilon$.

To give the lower bound, we estimate the number of choices $v_1 \in A \setminus (A_0 \cup A_1)$, $v_2 \in B_{v_1}$ and $v_3 \in C_{v_1}$ such that $v_2$ and $v_3$ are adjacent. We have $|A \setminus (A_0 \cup A_1)| \geq (1 - 2\delta)n$ and $e(B_{v_1}, C_{v_1}) \geq (d(B, C) - \varepsilon)|B_{v_1}||C_{v_1}| \geq (d(B, C) - \varepsilon)(d(A, B) - \varepsilon)(d(A, C) - \varepsilon)n^2$. Hence, the number of triangles is at least

$$(1 - 2\delta)(d(B, C) - \varepsilon)(d(A, B) - \varepsilon)(d(A, C) - \varepsilon)n^3.$$

□

## 2 Forumlation of Szemeredi's regularity lemma

**Definition 2.** *For a graph $G$, a partition $V_0$, $V_1$, …, $V_m$ of $V(G)$ is* $\varepsilon$-regular *if*

- $|V_0| \leq \varepsilon |V(G)|$,
- $|V_1| = |V_2| = \ldots = |V_m|$, *and*
- *for all but at most $\varepsilon m^2$ values of $1 \leq i < j \leq m$, the pair $(V_i, V_j)$ is $\varepsilon$-regular.*

*The integer $m$ is called the* order *of the partition.*

**Theorem 7.** *For any positive integer $m_0$ and real number $\varepsilon > 0$, there exists an integer $M \geq m_0$ such that the following holds. Every graph $G$ with at least $m_0$ vertices has an $\varepsilon$-regular partition of order at least $m_0$ and at most $M$.*

# 3 Exercises

1. ($\star\star$) Let $m_0$ be an arbitrary positive integer and let $\varepsilon > 0$ be a real number. Show that there exists $n_0$ as follows. Let $G$ be any graph with at least $n_0$ vertices and at most $100|V(G)|$ edges. Let $A, B \subseteq V(G)$ be disjoint. If $|A|, |B| \geq |V(G)|/m_0$, then $(A, B)$ is an $\varepsilon$-regular pair in $G$.

2. ($\star\star$) Suppose that $(A, B)$ is a $(\delta, \varepsilon)$-regular pair in $G$ for some $0 < \delta, \varepsilon \leq 1$. Let $A_1 \subseteq A$ and $B_1 \subseteq B$ satisfy $|A_1| \geq \alpha|A|$ and $|B_1| \geq \alpha|B|$ for some $\alpha > 0$. Prove that $(A_1, B_1)$ is a $(\delta/\alpha, 2\varepsilon)$-regular pair.

3. ($\star\star$) Let $G$ be a graph and let $(A, B)$ be a $(\delta, \varepsilon)$-regular pair in $G$ for some $0 < \delta, \varepsilon \leq 1$, with $|A| = |B| = n$. Prove that the number of 4-cycles $v_1 v_2 v_3 v_4$ of $G$ with $v_1, v_3 \in A$ and $v_2, v_4 \in B$ is at most

$$\left[(d(A, B) + \varepsilon)^4 + 2\delta\right] n^4$$

(or give any similar upper bound roughly equal to $d(A, B)^4 n^4$).